

09/967,228

REMARKS

In view of the following discussion, the Applicants submit that none of the claims now pending in the application is anticipated under the provisions of 35 U.S.C. § 102. Thus, the Applicants believe that all of these claims are now in allowable form.

I. THE REJECTION OF CLAIMS 1-4 AND 7-10 UNDER 35 U.S.C. 102

The Examiner rejected claims 1-4 and 7-10 under 35 U.S.C. §102(e) as being anticipated by the Kuhn patent (United States Patent No. 6,029,132, issued February 22, 2000, hereinafter "Kuhn"). The Applicants respectfully traverse the rejection.

Kuhn teaches a method for generating audible pronunciations for text strings (e.g., spelled words). A text string, such as a sentence, is received by a text-based pronunciation generator, which employs a set of decision trees that aid in generating a list of potential pronunciations for the input text string. The decision trees base pronunciation determinations on several factors including the syntaxes or parts of speech of each word in the input text string. An estimator employing a set of phoneme-mixed decision trees is then used to assign a probability score to each potential pronunciation by sequentially examining each letter in the input text string, along with phonemes assigned to each letter by the pronunciation generator. In this way, more accurate pronunciations of input text strings are produced.

The Examiner's attention is directed to the fact that Kuhn does not teach, show or suggest the novel invention of applying recognition passes to a speech signal in order to recognize an utterance containing the speech signal, as positively recited by the Applicants' independent claims 1, 7, 9 and 10. Specifically, claims 1, 7, 9 and 10 recite:

1. A method for recognizing an utterance that pertains to a sparse domain, the sparse domain having a linguistic structure and a plurality of components, objects or concepts, the method comprising the steps of:

acquiring a speech signal that represents an utterance;

performing a first recognition pass by applying a first language model to the speech signal;

selecting or generating a second language model based at least in part on results from the first recognition pass, on information regarding a linguistic structure of a

09/967,228

domain within the speech signal, and on information regarding relationships among the domain components, objects or concepts within the speech signal; and

performing a second recognition pass by applying the second language model to at least a portion of the speech signal to recognize the utterance containing the speech signal. (Emphasis added)

7. In a speech recognition system, a method for recognizing an utterance comprising the steps of:

acquiring a speech signal that represents the utterance; and

performing a series of recognition passes, a second and subsequent recognition passes processing at least a portion of the speech signal using a language model that is constrained by a result of a previous recognition pass. (Emphasis added)

9. A method for generating language models between speech recognition passes, the language models based on a domain having a linguistic structure and a plurality of components, objects or concepts, the method comprising the steps of:

generating or acquiring a database containing information regarding the linguistic structure of the domain and information regarding relationships among the domain components, objects or concepts;

acquiring a result from a speech recognition pass, the result including a domain component, object or concept; and

generating a language model that includes a subset of the domain by using the result from the speech recognition pass to select information from the database. (Emphasis added)

10. In a speech recognition system, a method for generating language models based on a domain having a plurality of components, objects or concepts, the method comprising the steps of:

acquiring a result from a speech recognition pass, the result including a domain component, object or concept;

using the result from the speech recognition pass to perform a search on a database that contains information regarding relationships among the domain components, objects or concepts; and

generating a language model using a result from the database search. (Emphasis added)

The Applicants' invention is directed to a method and apparatus for performing relational speech recognition. In various speech recognition applications, including speech recognition-based Global Positioning System (GPS) navigation systems used in automobiles, it is particularly desirable to produce accurate speech recognition results

09/967,228

based on limited user input. Existing applications typically require a great deal of user interaction in order to produce the correct results; for example an application may ask a user to speak only one word at a time for recognition, or may ask a user several questions after performing an initial recognition pass in order to further refine the results. This is not only time consuming for the user, but may distract the user from other tasks, such as driving.

The Applicants' invention addresses the aforementioned concerns by making use of observable relationships among words in sparse domains. A first recognition pass is performed by applying a first language model to a query (e.g., an acquired speech signal) in order to recognize at least some of the words in the query. These results may then be combined with information regarding the linguistic structure of the query's domain, or information regarding relationships among concepts, objects or components in the query's domain, in order to form a second language model having a more narrowly tailored search space than the first language model. This second language model is then applied to at least a portion of the original query in order to refine the results obtained by the first recognition pass using the first language model. This iterative process may be repeated several times until satisfactory recognition of the spoken query is achieved.

By contrast, Kuhn only teaches acquiring a text-based input (e.g., a sentence or a string) and applying several decision trees to the text-based input in order to generate a pronunciation for audible or spoken output representing the text-based input. Kuhn does not teach the reverse application, e.g., acquiring an audible or spoken input (e.g., a user query) and identifying words in the audible input in order to produce recognized speech, for example in the form of a text-based transcription of the spoken input. Although Kuhn discloses that the generated pronunciations may be used to create a library for use in speech-to-text applications, Kuhn does not disclose a method by which speech may be converted to text using a library generated in this way. Thus, for at least these reasons, Kuhn fails to anticipate the invention recited by claims 1, 7, 9 and 10, and claims 1, 7, 9 and 10 are therefore patentable over Kuhn.

Claims 2-4 and 8 depend, respectively, from independent claims 1 and 7 and

09/967,228

recite additional limitations therefore. Thus, for at least the reasons stated above, claims 2-4 and 8 are also not anticipated by Kuhn and are patentable under 35 U.S.C. §102(e). Accordingly, the Applicants respectfully request that the rejection of claims 1-4 and 7-10 over Kuhn be withdrawn.

II. THE REJECTION OF CLAIMS 5-6 AND 11-17 UNDER 35 U.S.C. 102

The Examiner rejected claims 5-6 and 11-17 under 35 U.S.C. §102(e) as being anticipated by the Junqua patent (United States Patent No. 6,314,165, issued February November 6, 2001, hereinafter "Junqua"). The Applicants respectfully traverse the rejection.

Junqua teaches an automated hotel attendant for routing phone calls to specified hotel guests using speech recognition. The attendant includes a lexicon training system that generates audible or spoken pronunciations from the text form of a guest's name. Specifically, the lexicon training system employs a set of decision trees that aid in generating a list of potential pronunciations for the input text. The decision trees base pronunciation determinations in part on the letters that form the guest's name. An estimator employing a set of phoneme-mixed decision trees is then used to assign a probability score to each potential pronunciation by sequentially examining each letter in the input text, along with phonemes assigned to each letter. Once a proper pronunciation for the guest's name is generated, it may be used to train a speech recognition system so that a user speaking the guest's name will be connected to the guest's extension.

The Examiner's attention is directed to the fact that Junqua does not teach, show or suggest the novel invention of applying recognition passes to a speech signal in order to recognize an utterance containing the speech signal, as positively recited by the Applicants' independent claims 5 and 11. Specifically, claims 5 and 11 recite:

5. A method for recognizing an utterance pertaining to an address or location, each address or location having a plurality of components, the method comprising the steps of:

acquiring a speech signal that represents an utterance;
performing a first recognition pass by applying a first language model to the

09/967,228

speech signal;

selecting or generating a second language model based at least in part on results from the first recognition pass and on information regarding relationships among the address or location components; and

performing a second recognition pass by applying the second language model to at least a portion of the speech signal to recognize the utterance contained in the speech signal. (Emphasis added)

11. A method for recognizing an address or location expressed as a single utterance, the method comprising the steps of:

acquiring a speech signal that represents the single utterance; and

performing a series of recognition passes, a second and subsequent recognition passes processing at least a portion of the speech signal using a language model that is constrained by a result of a previous recognition pass. (Emphasis added)

As discussed above, the Applicants' invention is directed to a method and apparatus for performing relational speech recognition. A first recognition pass is performed by applying a first language model to a spoken query, such as an address, in order to recognize at least some of the words in the query. These results may then be used to generate a second language model having a more narrowly tailored search space than the first language model. This second language model is then applied to at least a portion of the original query in order to refine the results obtained by the first recognition pass using the first language model. This iterative process may be repeated several times until satisfactory recognition of the spoken query is achieved.

By contrast, Junqua only teaches acquiring a text-based input (e.g., a guest's name) and applying several decision trees to the text-based input in order to generate a pronunciation for audible or spoken output representing the text-based input. Junqua does not teach the reverse application, e.g., acquiring an audible or spoken input (e.g., a user query) and performing recognition passes on the audible input in order to produce recognized speech, for example in the form of a text-based transcription of the spoken input. Although Junqua generically discloses that the generated pronunciations may be used to create a library for use in speech recognition applications, Junqua does not disclose an actual method by which the speech may be recognized, and certainly does not teach applying one or more language models to an acquired speech signal in

09/967,228

order recognize an utterance contained therein. Thus, for at least these reasons, Junqua fails to anticipate the invention recited by claims 5 and 11, and claims 5 and 11 are therefore patentable over Junqua.

Claims 6 and 12-17 depend, respectively, from independent claims 5 and 11 and recite additional limitations therefore. Thus, for at least the reasons stated above, claims 6 and 12-17 are also not anticipated by Junqua and are patentable under 35 U.S.C. §102(e). Accordingly, the Applicants respectfully request that the rejection of claims 6 and 12-17 over Junqua be withdrawn.

Conclusion

Thus, the Applicants submit that all of the presented claims fully satisfy the requirements of 35 U.S.C. §102. Consequently, the Applicants believe that all these claims are presently in condition for allowance. Accordingly, both reconsideration of this application and its swift passage to issue are earnestly solicited.

If, however, the Examiner believes that there are any unresolved issues requiring the issuance of a final action in any of the claims now pending in the application, it is requested that the Examiner telephone Mr. Kin-Wah Tong, Esq. at (732) 530-9404 so that appropriate arrangements can be made for resolving such issues as expeditiously as possible.

9/7/04
Date

Respectfully submitted,



Kin-Wah Tong, Attorney
Reg. No. 39,400
(732) 530-9404

Moser, Patterson & Sheridan, LLP
595 Shrewsbury Avenue
Shrewsbury, New Jersey 07702